Classification With A Preassigned Error Rate When Two Covariance Matrices Are Equal

by

Muni S. Srivastava
Department of Statistics
University of Toronto

and

Makoto Aoshima
Department of Mathematics and Informatics
Tokyo Gakugei University

Technical Report No. 9807, March 13, (1998)

TECHNICAL REPORT SERIES

University of Toronto
Department of Statistics

CLASSIFICATION WITH A PREASSIGNED ERROR RATE WHEN TWO COVARIANCE MATRICES ARE EQUAL

Muni S. Srivastava
Department of Statistics
University of Toronto
Toronto, Ontario
Canada M5S 3G3

Makoto Aoshima
Department of Mathematics and Informatics
Tokyo Gakugei University
Tokyo 184, Japan

Key words and phrases: Asymptotic efficiency; discrimination rule; exact consistency; Monte Carlo simulation; probability requirement; two-stage procedure.

ABSTRACT

We consider the problem of constructing a discrimination rule which controls the two probabilities of misclassification at a specified level for classifying an individual into two p-variate normal populations having a common and unknown covariance matrix. A two-stage procedure is proposed to develop such a discrimination rule. It is shown that the proposed two-stage procedure is asymptotically efficient. The efficiency of the proposed two-stage procedure is demonstrated by Monte Carlo simulation.

1. INTRODUCTION

Suppose that there are three populations π_i, $i = 0, 1, 2$, where it is known that $\pi_0 = \pi_i$ for exactly one of $i = 1, 2$, but we do not know for which i. The problem is to find for which i this is true. In the investigation it is assumed that π_i's are independently distributed as p-variate normal distributions $N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$, $i = 0, 1, 2$, where all the parameters are unknown and $\boldsymbol{\Sigma} > \boldsymbol{O}$. The problem is, thus, reduced to finding the population with $\boldsymbol{\mu}_i = \boldsymbol{\mu}_0$. For this problem we consider a discrimination rule which satisfies the requirement that

$$\max(e_{12}, e_{21}) \leq \alpha \quad \text{whenever} \quad \boldsymbol{\delta}'\boldsymbol{\delta} \geq d^2 \tag{1.1}$$

exactly for specified constants $d\ (> 0)$ and $\alpha\ (0 < \alpha < 1/2)$, where $\boldsymbol{\delta} = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$, and e_{ij} $(i, j = 1, 2,\ i \neq j)$ denotes the probability of misclassifying π_0 from π_i into π_j. It should be noted that the requirement (1.1) needs the samples drawn from π_1 and π_2 to which π_0 is to be classified to be of the same sample size. Although it can be extended to the unequal sample size case by requiring that $e_{12} \leq \alpha$ and $e_{21} \leq \beta$, we consider only the equal sample size case in this paper.

Since when $\boldsymbol{\Sigma}$ is unknown, there does not exist a fixed-sample size procedure which controls the two probabilities of misclassification at a specified level, Srivastava (1973) proposed a purely sequential procedure and showed that the associated discrimination rule satisfies the requirement (1.1) asymptotically when $d \to 0$. Aoshima and Aoki (1997) transformed Srivastava's purely sequential procedure into the two-stage procedure which has the same asymptotic efficiency as Srivastava's procedure. Recently, Aoshima and Shimono (1998) gave a two-stage discrimination rule for $p = 1$ which satisfies the requirement (1.1) exactly. In this paper we shall develop a multivariate two-stage procedure for this problem which gives a discrimination rule satisfying the requirement (1.1) exactly for the multivariate case. Then, the requirement (1.1) puts the restriction that the sample sizes drawn from π_1 and π_2 must be the same. Although the derivation of the proposed procedure here is quite different from the case $p = 1$, it includes the result for $p = 1$ as a special case since the errors of misclassification are decreasing function of the noncentrality parameter. The outline of this paper is as follows:

In Section 2, we give a fixed-sample size procedure for the case when Σ were known. The discrimination rule is based on the maximum likelihood method and satisfies the requirement (1.1) exactly. The samples are, however, chosen to minimize the total sample size. In Section 3, a two-stage procedure is proposed by estimating the optimal fixed-sample size and it is shown that the associated discrimination rule satisfies the requirement (1.1) exactly for all $(\boldsymbol{\delta}, \Sigma, d, \alpha)$. In Section 4, some asymptotic properties of the proposed two-stage procedure is discussed and its efficiency is compared with Aoshima and Aoki's (1997) procedure which has the same efficiency as Srivastava's (1973) procedure asymptotically. Finally, in Section 5, a Monte Carlo simulation is given to illustrate the efficiency of the proposed two-stage procedure.

2. OPTIMAL FIXED-SAMPLE SIZE

When Σ were known, $\boldsymbol{\mu}_i$, $i = 0, 1, 2$, are estimated by the sample mean vectors $\overline{\boldsymbol{x}}_{in_i} = \sum_{j=1}^{n_i} \boldsymbol{x}_{ij}/n_i$, $i = 0, 1, 2$, where the random samples $\boldsymbol{x}_{i1}, ..., \boldsymbol{x}_{in_i}$ of size n_i are taken from π_i, $i = 0, 1, 2$, and $n_1 = n_2$ $(\equiv n)$. Then a discrimination rule based on the maximum likelihood method is as follows: The population π_0 is classified into π_1 if

$$(\overline{\boldsymbol{x}}_{0n_0} - \overline{\boldsymbol{x}}_{1n})' \Sigma^{-1} (\overline{\boldsymbol{x}}_{0n_0} - \overline{\boldsymbol{x}}_{1n}) - (\overline{\boldsymbol{x}}_{0n_0} - \overline{\boldsymbol{x}}_{2n})' \Sigma^{-1} (\overline{\boldsymbol{x}}_{0n_0} - \overline{\boldsymbol{x}}_{2n}) \leq 0, \quad (2.1)$$

and into π_2 otherwise (Srivastava and Khatri (1979), p.245.) Let $\boldsymbol{y}_1 = c\Sigma^{-1/2}$ $(\overline{\boldsymbol{x}}_{0n_0} - \overline{\boldsymbol{x}}_{1n})$ and $\boldsymbol{y}_2 = c\Sigma^{-1/2}(\overline{\boldsymbol{x}}_{0n_0} - \overline{\boldsymbol{x}}_{2n})$, where $c = (1/n_0 + 1/n)^{-1/2}$. Then $\boldsymbol{y}_1 - \boldsymbol{y}_2$ and $\boldsymbol{y}_1 + \boldsymbol{y}_2$ are independently distributed as $N_p(-c\boldsymbol{\delta}^*, 2(1-\rho)\boldsymbol{I}_p)$ and $N_p(c\boldsymbol{\delta}^*, 2(1+\rho)\boldsymbol{I}_p)$ when $\pi_0 = \pi_1$ and are independently distributed as $N_p(-c\boldsymbol{\delta}^*, 2(1-\rho)\boldsymbol{I}_p)$ and $N_p(-c\boldsymbol{\delta}^*, 2(1+\rho)\boldsymbol{I}_p)$ when $\pi_0 = \pi_2$, where $\rho = c^2/n_0$ and $\boldsymbol{\delta}^* = \Sigma^{-1/2}\boldsymbol{\delta}$. Let $\boldsymbol{u} = \boldsymbol{\Gamma}(\boldsymbol{y}_1 - \boldsymbol{y}_2)/\sqrt{2(1-\rho)}$ and $\boldsymbol{v} = \boldsymbol{\Gamma}(\boldsymbol{y}_1 + \boldsymbol{y}_2)/\sqrt{2(1+\rho)}$, where

$$\boldsymbol{\Gamma} = \begin{pmatrix} \boldsymbol{\delta}^{*'}/\Delta \\ \boldsymbol{B}' \end{pmatrix} \quad (2.2)$$

is an orthogonal matrix and $\Delta = \sqrt{\boldsymbol{\delta}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\delta}}$. Then $\boldsymbol{u} = (u_1, \boldsymbol{u}_2')'$ and $\boldsymbol{v} = (v_1, \boldsymbol{v}_2')'$ are independently distributed as

$$N_p\left((-c\Delta/\sqrt{2(1-\rho)}, \mathbf{o}')', \boldsymbol{I}_p\right) \text{ and } N_p\left((c\Delta/\sqrt{2(1+\rho)}, \mathbf{o}')', \boldsymbol{I}_p\right)$$

when $\pi_0 = \pi_1$ and are independently distributed as

$$N_p\left((-c\Delta/\sqrt{2(1-\rho)}, \mathbf{o}')', \boldsymbol{I}_p\right) \text{ and } N_p\left((-c\Delta/\sqrt{2(1+\rho)}, \mathbf{o}')', \boldsymbol{I}_p\right)$$

when $\pi_0 = \pi_2$. When $\pi_0 = \pi_1$, we have

$$\begin{aligned} e_{12} &= P\left(\boldsymbol{y}_1'\boldsymbol{y}_1 - \boldsymbol{y}_2'\boldsymbol{y}_2 > 0\right) \\ &= P\left(\boldsymbol{u}'\boldsymbol{v} > 0\right) \\ &= P\left(\boldsymbol{w}'\boldsymbol{v} - \frac{c\Delta}{\sqrt{2(1-\rho)}} v_1 > 0\right), \end{aligned}$$

where $\boldsymbol{w} = \left(u_1 + c\Delta/\sqrt{2(1-\rho)}, \boldsymbol{u}_2'\right)'$ is distributed as $N_p(\mathbf{o}, \boldsymbol{I}_p)$ being independent of $\boldsymbol{v}$. In this paper, we assume that $p \geq 2$. We note that when $p = 1$, the term containing $\boldsymbol{u}_2$ disappears and because of the monotonicity of the errors of misclassification, the results similar to $p \geq 2$ can be obtained. A separate treatment for $p = 1$ is, however, given in Aoshima and Shimono (1998). Thus, for $p \geq 2$,

$$e_{12} = 1 - E\left\{\Phi\left(\frac{c\Delta}{\sqrt{2(1-\rho)}} \frac{v_1}{\sqrt{v_1^2 + g}}\right)\right\},$$

where v_1 and $g = \boldsymbol{v}_2'\boldsymbol{v}_2$ are independently distributed as $N\left(c\Delta/\sqrt{2(1+\rho)}, 1\right)$ and χ^2_{p-1}, respectively, and $\Phi(\cdot)$ denotes the distribution function of a $N(0,1)$ random variable. Hence, we can write that

$$e_{12} = 1 - E\left\{\Phi\left(\frac{c\Delta}{\sqrt{2(p-1)(1-\rho)}} \frac{t}{\sqrt{1 + t^2/(p-1)}}\right)\right\}, \qquad (2.3)$$

where t denotes a noncentral t random variable with noncentrality parameter $\gamma = \sqrt{c^2\Delta^2/\{2(1+\rho)\}}$ and with $p-1$ degrees of freedom (d.f.). When $\pi_0 = \pi_2$, we have in a similar way that

$$\begin{aligned} e_{21} &= P\left(\boldsymbol{u}'\boldsymbol{v} \le 0\right) \\ &= 1 - E\left\{\Phi\left(\frac{c\Delta}{\sqrt{2(1-\rho)}}\frac{-v_1}{\sqrt{v_1^2+g}}\right)\right\}, \end{aligned}$$

where $-v_1$ and g are independently distributed as $N\left(c\Delta/\sqrt{2(1+\rho)}, 1\right)$ and χ^2_{p-1}, respectively. Hence, we have

$$e_{21} = 1 - E\left\{\Phi\left(\frac{c\Delta}{\sqrt{2(p-1)(1-\rho)}}\frac{t}{\sqrt{1+t^2/(p-1)}}\right)\right\}. \tag{2.4}$$

It is known that the errors of misclassification are monotone decreasing function of Δ (see Das Gupta (1974) and Srivastava and Khatri (1979)). Thus, when we let λ be the maximum latent root of $\boldsymbol{\Sigma}$, it holds that $\Delta \ge d/\sqrt{\lambda}$. Hence, we have for $i, j = 1, 2$ $(i \ne j)$ that

$$e_{ij} \le 1 - E\left\{\Phi\left(\frac{cd}{\sqrt{2(p-1)\lambda(1-\rho)}}\frac{\tilde{t}}{\sqrt{1+\tilde{t}^2/(p-1)}}\right)\right\}, \tag{2.5}$$

where $\tilde{t}$ denotes a noncentral t random variable with noncentrality parameter $\tilde{\gamma} = \sqrt{c^2d^2/\{2\lambda(1+\rho)\}}$ and with $p-1$ d.f.

Let us set $\lambda\left(1/n_0 + 1/n\right) = d^2/g_p^{*2}(\alpha)$ as in Aoshima and Shimono (1998), where $g^* = g_p^*(\alpha)$ is chosen as a solution to the equation

$$E\left\{\Phi\left(\frac{g^*}{\sqrt{2(p-1)(1-\rho)}}\frac{\tilde{t}}{\sqrt{1+\tilde{t}^2/(p-1)}}\right)\right\} = 1 - \alpha, \tag{2.6}$$

where $\tilde{t}$ is a noncentral t random variable with noncentrality parameter $\tilde{\gamma} = \sqrt{g^{*2}/\{2(1+\rho)\}}$ and with $p-1$ d.f. Then the discrimination rule (2.1) meets the requirement (1.1) exactly through (2.6). To determine the sample sizes n and n_0 in some optimum way, we choose ρ to minimize the total sample size $n_0 + 2n$. Then we have $\rho = \sqrt{2} - 1$. This gives

$$n_0 \ge \left(\sqrt{2}+1\right)\frac{g_p^{*2}(\alpha)}{d^2}\lambda \ (= n_0^*, \text{ say}) \tag{2.7}$$

for π_0, and

$$n \geq \frac{\left(\sqrt{2}+1\right)}{\sqrt{2}} \frac{g_p^{*2}(\alpha)}{d^2} \lambda \ (= n^*, \text{ say}) \tag{2.8}$$

for π_i, $i = 1, 2$. Note that the optimal fixed-sample sizes n_0^* and n^* defined above meet an asymptotically optimal choice given by Bechhofer and Turnbull (1971) for the one-sided comparisons problem with a control. (Cf. Hochberg and Tamhane (1987), p.202.)

Table 1 gives values of $g_p^*(\alpha)$ for $p = 2(1)5$ and $\alpha = .01$ and $.05$. These values were computed by solving the equation (2.6) via the bisection method. The expectation was computed by the Monte Carlo method with 10,000 independent trials.

TABLE 1 Values of $g_p^*(\alpha)$.

$\alpha \backslash p$	2	3	4	5
.01	4.18	4.27	4.39	4.50
.05	3.08	3.22	3.35	3.45

3. TWO-STAGE PROCEDURE

Since $\boldsymbol{\Sigma}$ is unknown, that is λ is unknown, the sample sizes n_0^* and n^* given by (2.7)-(2.8) required in the discrimination rule (2.1), are estimated by the following two-stage procedure.

We first take the initial samples $\boldsymbol{x}_{i1}, ..., \boldsymbol{x}_{im}$, of given size m ($\geq p/3 + 1$) from π_i, $i = 0, 1, 2$, and compute $\overline{\boldsymbol{x}}_{im} = \sum_{j=1}^{m} \boldsymbol{x}_{ij}/m$, $i = 0, 1, 2$, and

$$\boldsymbol{S}_m = \sum_{i=0}^{2} \sum_{j=1}^{m} (\boldsymbol{x}_{ij} - \overline{\boldsymbol{x}}_{im})(\boldsymbol{x}_{ij} - \overline{\boldsymbol{x}}_{im})'/\nu, \ \nu = 3(m-1). \tag{3.1}$$

Let ℓ_m be the maximum latent root of $\boldsymbol{S}_m$. The sample sizes of two-stage procedure are then defined by

$$N_0 = \max\left\{m, \left[(\sqrt{2}+1)\frac{g_{p,m}^2(\alpha)}{d^2}\ell_m\right] + 1\right\} \tag{3.2}$$

for π_0, and by

$$N_i = \max\left\{m, \left[\frac{(\sqrt{2}+1)}{\sqrt{2}}\frac{g_{p,m}^2(\alpha)}{d^2}\ell_m\right]+1\right\} \ (= N,\ \text{say}) \tag{3.3}$$

for π_i, $i = 1, 2$, where $[x]$ denotes the largest integer less than x. In the above definitions (3.2)-(3.3), the constant $g = g_{p,m}(\alpha)$ is chosen as a solution to the equation

$$E\left\{\Phi\left(\sqrt{\frac{\sqrt{2}+1}{2\sqrt{2}}}\frac{g}{\sqrt{\nu}}\frac{\tilde{v}/\sqrt{w_{1\cdot 2}}}{\sqrt{(\tilde{v}/w_{1\cdot 2})^2 + \tilde{\boldsymbol{h}}'\boldsymbol{W}_{22}^{-2}\tilde{\boldsymbol{h}}}}\right)\right\} = 1 - \alpha, \tag{3.4}$$

where $\tilde{\boldsymbol{h}} = \boldsymbol{t} - (\tilde{v}/w_{1\cdot 2})\boldsymbol{w}_{12}$, $\boldsymbol{t}$ is a $N_{p-1}(\mathbf{o}, \boldsymbol{I}_{p-1})$ random vector, $\tilde{v}$ is a $N\left(g\sqrt{1/(2\sqrt{2})}\sqrt{w_{1\cdot 2}/\nu} - \boldsymbol{w}_{12}'\boldsymbol{W}_{22}^{-1}\boldsymbol{t},\ 1\right)$ random variable, $w_{1\cdot 2} = w_{11} - \boldsymbol{w}_{12}'\boldsymbol{W}_{22}^{-1}\boldsymbol{w}_{12}$, and

$$\boldsymbol{W} = \begin{pmatrix} w_{11} & \boldsymbol{w}_{12}' \\ \boldsymbol{w}_{12} & \boldsymbol{W}_{22} \end{pmatrix}$$

denotes a $W_p(\nu, \boldsymbol{I}_p)$ matrix. Here, $\boldsymbol{t}$ and $\boldsymbol{W}$ are independent.

At the second stage we take the additional samples $\boldsymbol{x}_{im+1}, \ldots, \boldsymbol{x}_{iN_i}$, of size $N_i - m$ from π_i, $i = 0, 1, 2$, and compute $\overline{\boldsymbol{x}}_{iN_i} = \sum_{j=1}^{N_i}\boldsymbol{x}_{ij}/N_i$ to estimate $\boldsymbol{\mu}_i$ $(i = 0, 1, 2)$. Then the following discrimination rule is proposed: The population π_0 is classified into π_1 if

$$(\overline{\boldsymbol{x}}_{0N_0} - \overline{\boldsymbol{x}}_{1N})'\boldsymbol{S}_m^{-1}(\overline{\boldsymbol{x}}_{0N_0} - \overline{\boldsymbol{x}}_{1N}) - (\overline{\boldsymbol{x}}_{0N_0} - \overline{\boldsymbol{x}}_{2N})'\boldsymbol{S}_m^{-1}(\overline{\boldsymbol{x}}_{0N_0} - \overline{\boldsymbol{x}}_{2N}) \leq 0, \tag{3.5}$$

and into π_2 otherwise.

For the discrimination rule (3.5) we have

THEOREM 1 *The discrimination rule* (3.5) *satisfies the requirement* (1.1) *exactly.*

In the proof of this theorem the following lemma is useful.

LEMMA 1 *Let* $\boldsymbol{W}$ *be a* $p \times p$ *symmetric nonsingular matrix and*

$$\boldsymbol{W} = \begin{pmatrix} w_{11} & \boldsymbol{w}_{12}' \\ \boldsymbol{w}_{12} & \boldsymbol{W}_{22} \end{pmatrix},$$

where w_{11} is a scalar, $\boldsymbol{w}_{12}$ is a $(p-1)\times 1$ vector, and $\boldsymbol{W}_{22}$ is a $(p-1)\times(p-1)$ matrix. Let $w_{1\cdot 2} = w_{11} - \boldsymbol{w}_{12}'\boldsymbol{W}_{22}^{-1}\boldsymbol{w}_{12}$. Then

$$\boldsymbol{W}^{-1} = \begin{pmatrix} 0 & \mathbf{o}' \\ \mathbf{o} & \boldsymbol{W}_{22}^{-1} \end{pmatrix} + w_{1\cdot 2}^{-1} \begin{pmatrix} 1 \\ -\boldsymbol{W}_{22}^{-1}\boldsymbol{w}_{12} \end{pmatrix} \begin{pmatrix} 1, & -\boldsymbol{w}_{12}'\boldsymbol{W}_{22}^{-1} \end{pmatrix}.$$

PROOF See Srivastava and Khatri (1979, Corollary 1.4.2, p.8). ■

PROOF OF THEOREM 1 Let $\boldsymbol{\delta}^\star = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\delta}$, $\boldsymbol{y} = \tilde{c}_1\boldsymbol{\Sigma}^{-1/2}(2\overline{\boldsymbol{x}}_{0N_0} - \overline{\boldsymbol{x}}_{1N} - \overline{\boldsymbol{x}}_{2N})$ and $\boldsymbol{z} = \tilde{c}_2\boldsymbol{\Sigma}^{-1/2}(\overline{\boldsymbol{x}}_{2N} - \overline{\boldsymbol{x}}_{1N})$, where $\tilde{c}_1 = (4/N_0 + 2/N)^{-1/2}$ and $\tilde{c}_2 = (2/N)^{-1/2}$. Then, for given $\boldsymbol{S}_m$, $\boldsymbol{y}$ is distributed as $N_p(\tilde{c}_1\boldsymbol{\delta}^\star, \boldsymbol{I}_p)$ when $\pi_0 = \pi_1$ and as $N_p(-\tilde{c}_1\boldsymbol{\delta}^\star, \boldsymbol{I}_p)$ when $\pi_0 = \pi_2$, and $\boldsymbol{z}$ is distributed as $N_p(-\tilde{c}_2\boldsymbol{\delta}^\star, \boldsymbol{I}_p)$. Let $\boldsymbol{W} = \nu\boldsymbol{\Sigma}^{-1/2}\boldsymbol{S}_m\boldsymbol{\Sigma}^{-1/2}$. Then $\boldsymbol{W}$ is distributed as $W_p(\nu, \boldsymbol{I}_p)$. Note that we have $e_{12} = P(\boldsymbol{y}'\boldsymbol{W}^{-1}\boldsymbol{z} > 0)$ when $\pi_0 = \pi_1$, and $e_{21} = P\left((-\boldsymbol{y})'\boldsymbol{W}^{-1}\boldsymbol{z} > 0\right)$ when $\pi_0 = \pi_2$. That is $e_{12} = e_{21}$, so that we consider the case when $\pi_0 = \pi_1$ hereafter.

Let $\tilde{\boldsymbol{y}} = \boldsymbol{\Gamma}\boldsymbol{y} = (\tilde{y}_1, \tilde{\boldsymbol{y}}_2')'$, $\tilde{\boldsymbol{z}} = \boldsymbol{\Gamma}\boldsymbol{z} = (\tilde{z}_1, \tilde{\boldsymbol{z}}_2')'$ and $\widetilde{\boldsymbol{W}} = \boldsymbol{\Gamma}\boldsymbol{W}\boldsymbol{\Gamma}'$, where $\boldsymbol{\Gamma}$ is the same as in (2.2). Then, for given $\boldsymbol{S}_m$, $\tilde{\boldsymbol{y}}$ and $\tilde{\boldsymbol{z}}$ are independently distributed as $N_p\left((\tilde{c}_1\Delta, \mathbf{o}')', \boldsymbol{I}_p\right)$ and $N_p\left((-\tilde{c}_2\Delta, \mathbf{o}')', \boldsymbol{I}_p\right)$, respectively. We also have

$$\begin{aligned} \widetilde{\boldsymbol{W}}^{-1} &\equiv \begin{pmatrix} \tilde{w}_{11} & \widetilde{\boldsymbol{w}}_{12}' \\ \widetilde{\boldsymbol{w}}_{12} & \widetilde{\boldsymbol{W}}_{22} \end{pmatrix}^{-1} \\ &= \begin{pmatrix} \boldsymbol{\delta}^{\star\prime}\boldsymbol{W}^{-1}\boldsymbol{\delta}^\star/\Delta^2 & \boldsymbol{\delta}^{\star\prime}\boldsymbol{W}^{-1}\boldsymbol{B}/\Delta \\ \boldsymbol{B}'\boldsymbol{W}^{-1}\boldsymbol{\delta}^\star/\Delta & \boldsymbol{B}'\boldsymbol{W}^{-1}\boldsymbol{B} \end{pmatrix}. \end{aligned} \tag{3.6}$$

Let $\tilde{w}_{1\cdot 2} = \tilde{w}_{11} - \widetilde{\boldsymbol{w}}_{12}'\widetilde{\boldsymbol{W}}_{22}^{-1}\widetilde{\boldsymbol{w}}_{12}$, $\tilde{u} = \tilde{y}_1 - \widetilde{\boldsymbol{w}}_{12}'\widetilde{\boldsymbol{W}}_{22}^{-1}\tilde{\boldsymbol{y}}_2$ and $\boldsymbol{h} = \tilde{\boldsymbol{y}}_2 - (\tilde{u}/\tilde{w}_{1\cdot 2})\widetilde{\boldsymbol{w}}_{12}$. Then we have from Lemma 1 that

$$\begin{aligned} \boldsymbol{y}'\boldsymbol{W}^{-1}\boldsymbol{z} &= \tilde{\boldsymbol{y}}'\widetilde{\boldsymbol{W}}^{-1}\tilde{\boldsymbol{z}} \\ &= \frac{\tilde{u}}{\tilde{w}_{1\cdot 2}}\tilde{z}_1 + \boldsymbol{h}'\widetilde{\boldsymbol{W}}_{22}^{-1}\tilde{\boldsymbol{z}}_2. \end{aligned}$$

Since $\boldsymbol{y}'\boldsymbol{W}^{-1}\boldsymbol{z}$ is distributed as $N\left(-(\tilde{u}/\tilde{w}_{1\cdot 2})\tilde{c}_2\Delta,\ (\tilde{u}/\tilde{w}_{1\cdot 2})^2 + \boldsymbol{h}'\widetilde{\boldsymbol{W}}_{22}^{-2}\boldsymbol{h}\right)$ for given $\tilde{\boldsymbol{y}}$ and $\widetilde{\boldsymbol{W}}$, we obtain that

$$e_{12} = e_{21} = 1 - E\left\{\Phi\left(\frac{(\tilde{u}/\tilde{w}_{1\cdot 2})\tilde{c}_2\Delta}{\sqrt{(\tilde{u}/\tilde{w}_{1\cdot 2})^2 + \boldsymbol{h}'\widetilde{\boldsymbol{W}}_{22}^{-2}\boldsymbol{h}}}\right)\right\}.$$

Now $\tilde{u}$ is distributed as $N(\tilde{c}_1\Delta - \widetilde{\boldsymbol{w}}'_{12}\widetilde{\boldsymbol{W}}_{22}^{-1}\widetilde{\boldsymbol{y}}_2,\ 1)$ for given $\widetilde{\boldsymbol{y}}_2$ and $\widetilde{\boldsymbol{W}}$. Noting that we can show that $e_{12}(= e_{21})$ is a decreasing function of Δ in a way similar to Das Gupta (1974). (See also Srivastava and Khatri (1979).) The inequality $\Delta \geq \sqrt{\widetilde{w}_{1\cdot 2}/\nu}\sqrt{d^2/\ell_m}$ holds in view of (3.6) and Lemma 1. Then we have

$$e_{12} = e_{21} \leq 1 - E\left\{\Phi\left(\frac{1}{\sqrt{\nu}}\frac{\left(\tilde{v}_1/\sqrt{\widetilde{w}_{1\cdot 2}}\right)\tilde{c}_2\sqrt{d^2/\ell_m}}{\sqrt{(\tilde{v}_1/\widetilde{w}_{1\cdot 2})^2 + \widetilde{\boldsymbol{h}}_1'\widetilde{\boldsymbol{W}}_{22}^{-2}\widetilde{\boldsymbol{h}}_1}}\right)\right\}, \tag{3.7}$$

where $\widetilde{\boldsymbol{h}}_1 = \widetilde{\boldsymbol{y}}_2 - (\tilde{v}_1/\widetilde{w}_{1\cdot 2})\widetilde{\boldsymbol{w}}_{12}$ and $\tilde{v}_1$ is distributed as $N\left(\tilde{c}_1\sqrt{\widetilde{w}_{1\cdot 2}/\nu}\sqrt{d^2/\ell_m} - \widetilde{\boldsymbol{w}}'_{12}\widetilde{\boldsymbol{W}}_{22}^{-1}\widetilde{\boldsymbol{y}}_2,\ 1\right)$. We also note that $e_{12}(= e_{21})$ is a decreasing function of $\tilde{c}_1$ and $\tilde{c}_2$. From (3.2)-(3.3) we obtain that

$$\tilde{c}_1 \geq \sqrt{\frac{1}{2\sqrt{2}}}\frac{\sqrt{\ell_m}}{d}g_{p,m}(\alpha) \ \text{ and } \ \tilde{c}_2 \geq \sqrt{\frac{\sqrt{2}+1}{2\sqrt{2}}}\frac{\sqrt{\ell_m}}{d}g_{p,m}(\alpha).$$

Therefore, we conclude from (3.7) that

$$e_{12} = e_{21} \leq 1 - E\left\{\Phi\left(\sqrt{\frac{\sqrt{2}+1}{2\sqrt{2}}}\frac{g_{p,m}(\alpha)}{\sqrt{\nu}}\frac{\tilde{v}_2/\sqrt{\widetilde{w}_{1\cdot 2}}}{\sqrt{(\tilde{v}_2/\widetilde{w}_{1\cdot 2})^2 + \widetilde{\boldsymbol{h}}_2'\widetilde{\boldsymbol{W}}_{22}^{-2}\widetilde{\boldsymbol{h}}_2}}\right)\right\},$$

where $\widetilde{\boldsymbol{h}}_2 = \widetilde{\boldsymbol{y}}_2 - (\tilde{v}_2/\widetilde{w}_{1\cdot 2})\widetilde{\boldsymbol{w}}_{12}$ and $\tilde{v}_2$ is distributed as $N\left(g_{p,m}(\alpha)\sqrt{1/(2\sqrt{2})}\sqrt{\widetilde{w}_{1\cdot 2}/\nu} - \widetilde{\boldsymbol{w}}'_{12}\widetilde{\boldsymbol{W}}_{22}^{-1}\widetilde{\boldsymbol{y}}_2,\ 1\right)$. The proof is complete in view of the definition of $g_{p,m}(\alpha)$ in (3.4). ∎

The following Table 2 gives values of $g_{p,m}(\alpha)$ for $p = 2(1)5$ and $m = 5(5)30(10)50, 100$ when $\alpha = .01$ and $.05$. These values were computed by solving the equation (3.4) via bisection method. The expectation was computed by the Monte Carlo method with 10,000 independent trials.

TABLE 2 Values of $g_{p,m}(\alpha)$.

$\alpha = .01$

$p \setminus m$	5	10	15	20	30	40	50	100
2	5.18	4.53	4.38	4.27	4.23	4.19	4.20	4.18
3	5.94	4.93	4.63	4.55	4.40	4.40	4.35	4.30
4	7.03	5.16	4.82	4.73	4.58	4.55	4.51	4.48
5	8.00	5.51	5.12	4.99	4.82	4.65	4.69	4.60

$\alpha = .05$

$p \setminus m$	5	10	15	20	30	40	50	100
2	3.64	3.27	3.20	3.16	3.12	3.08	3.13	3.08
3	4.20	3.58	3.43	3.39	3.31	3.28	3.27	3.24
4	4.91	3.82	3.66	3.53	3.46	3.44	3.44	3.39
5	5.63	4.12	3.87	3.73	3.64	3.58	3.57	3.53

REMARK 1 From (3.4) we can see that $g_{p,m}(\alpha) \to g_p^*(\alpha)$ as $m \to \infty$.

4. ASYMPTOTIC PROPERTIES

For the two-stage procedure proposed in Section 3, some asymptotic properties are studied under the condition that $m = m(d)$:

$$m(d) \to \infty,\ d^2 m(d) \to 0 \text{ as } d \to 0. \tag{4.1}$$

THEOREM 2 *The two-stage procedure based on* (3.2)-(3.3) *is asymptotically efficient, i.e.*

$$\lim_{d \to 0} \frac{E(N_0 + 2N)}{n_0^* + 2n^*} = 1.$$

PROOF From (3.2)-(3.3) we have

$$(3+2\sqrt{2})\frac{g_{p,m}^2(\alpha)}{d^2}E(\ell_m) \le E(N_0+2N) \le 3m+(3+2\sqrt{2})\frac{g_{p,m}^2(\alpha)}{d^2}E(\ell_m). \tag{4.2}$$

Since $n_0^* + 2n^* = (3+2\sqrt{2})g_p^{*2}(\alpha)\lambda/d^2$ from (2.7)-(2.8), it holds that

$$\frac{g_{p,m}^2(\alpha)}{g_p^{*2}(\alpha)}\frac{E(\ell_m)}{\lambda} \le \frac{E(N_0+2N)}{n_0^*+2n^*} \le \frac{3md^2}{(3+2\sqrt{2})g_p^{*2}(\alpha)\lambda} + \frac{g_{p,m}^2(\alpha)}{g_p^{*2}(\alpha)}\frac{E(\ell_m)}{\lambda}. \tag{4.3}$$

Since $E(\ell_m) \to \lambda$ as $m \to \infty$, and $g_{p,m}(\alpha) \to g_p^*(\alpha)$ from Remark 1 as $m \to \infty$, we obtain the result by combining those and (4.3) under the condition (4.1).

■

REMARK 2 From the left hand side of (4.2), it appears that

$$E\left(N_0+2N\right) \geq n_0^* + 2n^*,$$

since $E(\ell_m) \geq \lambda$ and $g_{p,m}(\alpha) \geq g_p^*(\alpha)$ from Tables 1-2. It has not been possible to show theoretically that $g_{p,m}(\alpha) \geq g_p^*(\alpha)$.

Srivastava's (1973) purely sequential procedure is transformed into the following two-stage procedure with the same asymptotic efficiency: After computing the maximum latent root ℓ_m of $\boldsymbol{S}_m$ as in (3.1), the sample size of two-stage procedure is defined by

$$\widetilde{N} = \max\left\{m, \left[\frac{6\tilde{g}_{p,m}^2(\alpha)}{d^2}\ell_m\right]+1\right\} \tag{4.4}$$

for π_i, $i = 0,1,2$. The constant $\tilde{g} = \tilde{g}_{p,m}(\alpha)$ is given as a solution to the equation

$$E\left[\Phi\left(\tilde{g}\sqrt{\frac{\lambda_{p,m}}{\nu}}\right)\right] = 1-\alpha,$$

where $\lambda_{p,m}$ is the minimum latent root of a $W_p(\nu, \boldsymbol{I}_p)$ matrix and $\nu = 3(m-1)$. At the second stage the additional samples $\boldsymbol{x}_{im+1}, \dots, \boldsymbol{x}_{i\widetilde{N}}$, of size $\widetilde{N}-m$ are taken from π_i, $i = 0,1,2$, and $\overline{\boldsymbol{x}}_{i\widetilde{N}} = \sum_{j=1}^{\widetilde{N}} \boldsymbol{x}_{ij}/\widetilde{N}$ is computed as an estimator of $\boldsymbol{\mu}_i$ for each π_i. Then a discrimination rule is given as follows: The population π_0 is classified into π_1 if

$$(\overline{\boldsymbol{x}}_{0\widetilde{N}} - \overline{\boldsymbol{x}}_{1\widetilde{N}})'\boldsymbol{S}_m^{-1}(\overline{\boldsymbol{x}}_{0\widetilde{N}} - \overline{\boldsymbol{x}}_{1\widetilde{N}}) - (\overline{\boldsymbol{x}}_{0\widetilde{N}} - \overline{\boldsymbol{x}}_{2\widetilde{N}})'\boldsymbol{S}_m^{-1}(\overline{\boldsymbol{x}}_{0\widetilde{N}} - \overline{\boldsymbol{x}}_{2\widetilde{N}}) \leq 0, \tag{4.5}$$

and into π_2 otherwise. Srivastava's (1973) purely sequential procedure and this two-stage procedure satisfy the requirement (1.1) asymptotically when $d \to 0$. (See Aoshima and Aoki (1997).)

We compare the proposed two-stage procedure with this two-stage procedure. Under the condition (4.1), we can prove that

$$\lim_{d\to 0}\frac{E(\widetilde{N})}{E(N_0)} = \frac{6\tilde{g}_{p,m}^2(\alpha)}{(\sqrt{2}+1)g_{p,m}^2(\alpha)}$$

for π_0,

$$\lim_{d\to 0} \frac{E(\widetilde{N})}{E(N)} = \frac{6\sqrt{2}\tilde{g}^2_{p,m}(\alpha)}{(\sqrt{2}+1)g^2_{p,m}(\alpha)}$$

for π_1 and π_2, and

$$\lim_{d\to 0} \frac{E(3\widetilde{N})}{E(N_0+2N)} = \frac{18\tilde{g}^2_{p,m}(\alpha)}{(2\sqrt{2}+3)g^2_{p,m}(\alpha)},$$

by using the same technique as in Theorem 2. These limiting values are given in Tables 3-5 for $p = 2(1)5$ and $m = 10(5)20(10)50$ when $\alpha = .01$ and $.05$.

TABLE 3 Values of $\lim_{d\to 0} E(\widetilde{N})/E(N_0)$ for π_0.

$\alpha = .01$

$p\backslash m$	10	15	20	30	40	50
2	.95	.91	.92	.90	.87	.87
3	.98	.96	.92	.91	.88	.87
4	1.06	1.01	.95	.91	.88	.86
5	1.09	1.00	.96	.90	.88	.85

$\alpha = .05$

$p\backslash m$	10	15	20	30	40	50
2	.88	.83	.82	.81	.81	.80
3	.89	.86	.82	.79	.78	.76
4	.93	.85	.84	.79	.76	.73
5	.94	.85	.83	.77	.74	.71

TABLE 4 Values of $\lim_{d\to 0} E(\widetilde{N})/E(N)$ for π_1 and π_2.

$\alpha = .01$

$p\backslash m$	10	15	20	30	40	50
2	1.34	1.29	1.29	1.27	1.23	1.22
3	1.39	1.36	1.30	1.28	1.24	1.23
4	1.50	1.42	1.34	1.28	1.24	1.21
5	1.53	1.41	1.36	1.27	1.24	1.19

$\alpha = .05$

$p\backslash m$	10	15	20	30	40	50
2	1.24	1.16	1.16	1.14	1.14	1.13
3	1.25	1.21	1.15	1.12	1.10	1.08
4	1.31	1.20	1.18	1.11	1.07	1.03
5	1.32	1.20	1.17	1.08	1.04	1.00

TABLE 5 Values of $\lim_{d\to 0} E(3\widetilde{N})/E(N_0 + 2N)$.

$\alpha = .01$

$p\backslash m$	10	15	20	30	40	50
2	1.18	1.13	1.14	1.11	1.08	1.08
3	1.22	1.20	1.15	1.13	1.09	1.08
4	1.32	1.25	1.18	1.13	1.09	1.07
5	1.35	1.24	1.20	1.11	1.09	1.05

$\alpha = .05$

$p\backslash m$	10	15	20	30	40	50
2	1.09	1.02	1.02	1.01	1.00	.99
3	1.10	1.06	1.01	.99	.97	.95
4	1.15	1.06	1.04	.98	.94	.90
5	1.17	1.05	1.03	.95	.92	.88

We can observe from these tables that the two-stage procedure given in (4.4) will usually undersample from π_0 and oversample from π_i, $i = 1, 2$. When α and m are small for p, the proposed two-stage procedure is asymptotically more efficient than the two-stage procedure (4.4). We also note that Srivastava's (1973) purely sequential procedure and the two-stage procedure (4.4) do not always satisfy the requirement (1.1) exactly for small m. (See also Aoshima and Aoki (1997).) Therefore, the proposed two-stage procedure in which samples are drawn from π_0 and π_i, $i = 1, 2$, in the ratio of $\sqrt{2}$ to 1 should be preferable to other classification rules.

5. SIMULATION

For the proposed two-stage procedure (3.2)-(3.3) and the two-stage procedure (4.4), we evaluate the required sample size and the associated classification probability (i.e. $1 - \max(e_{12}, e_{21})$) by computing the average numbers of $k = 10{,}000$ independent trials via the Monte Carlo simulation. Let $p = 3$. The simulation was carried out with three populations π_i, $i = 0, 1, 2$, generated by three independent sequences of pseudo normal random numbers which have the mean vectors $\boldsymbol{\mu}_1 = (\text{d}+.01, 0, 0)'$, $\boldsymbol{\mu}_2 = (0, 0, 0)'$ and $\boldsymbol{\mu}_0 = \boldsymbol{\mu}_1$ or $\boldsymbol{\mu}_2$, and the covariance matrix $\boldsymbol{\Sigma} = \text{diag}(3, 1, 1)$. Let $n^{**} = n_0^* + 2n^*$. We set $\alpha = .05$ and $d = 1.346, .952, .777$. Then the pairs (n_0^*, n^*) of the optimal fixed-sample sizes are (41.4, 29.3), (82.8, 58.6) and (124.3, 87.9), so that $n^{**} = 100$, 200 and 300, respectively, since $g_p^*(\alpha) = 3.22$ from Table 1 with $(p, \alpha) = (3, .05)$. The initial sample sizes of the two-stage procedures were set as $m = 10, 15, 20, 30$. Then, with $(p, \alpha) = (3, .05)$, $\tilde{g}_{p,m}(\alpha) = 2.14$, 2.01, 1.94 and 1.87 from Table 1 in Aoshima and Aoki (1997) and $g_{p,m}(\alpha) = 3.58$, 3.43, 3.39 and 3.31 from Table 2, for $m = 10$, 15, 20 and 30. For a value of n^{**} and m, suppose that a sample size from each population has the observed value n_{ij}, $i = 0, 1, 2$, respectively, in the j^{th} replication, and also let $p_j = 1$ or 0 according as each discrimination rule, in the j^{th} replication, $j = 1, \ldots, k$. Let $\overline{N}_i = k^{-1} \sum_{j=1}^{k} n_{ij}$, $S(\overline{N}_i) = [\{k(k-1)\}^{-1} \sum_{j=1}^{k} (n_{ij} - \overline{N}_i)^2]^{1/2}$, $\overline{CP} = k^{-1} \sum_{j=1}^{k} p_i$ and $S(\overline{CP}) = [\{\overline{CP}(1 - \overline{CP})/k\}]^{1/2}$, $i = 0, 1, 2$.

TABLE 6 Comparison of the proposed procedure (3.2)-(3.3) and the procedure (4.4)

Proposed procedure (3.2)-(3.3) $(i = 1, 2)$					
m	n^{**}	$(\overline{N}_0,\ S(\overline{N}_0))$	$(\overline{N}_i,\ S(\overline{N}_i))$	$\sum_{i=0}^{2} \overline{N}_i$	$(\overline{CP},\ S(\overline{CP}))$
10	100	(54.0, .374)	(38.3, .314)	130.6	(.980, .00139)
	200	(106.9, .524)	(75.7, .440)	258.3	(.980, .00139)
	300	(161.2, .647)	(114.1, .544)	389.5	(.981, .00137)
15	100	(48.4, .322)	(34.4, .271)	117.2	(.975, .00156)
	200	(97.2, .459)	(68.9, .386)	234.9	(.975, .00158)
	300	(145.6, .552)	(103.1, .464)	351.9	(.979, .00142)
20	100	(47.2, .292)	(33.6, .246)	114.3	(.972, .00164)
	200	(94.3, .417)	(66.8, .351)	228.0	(.975, .00155)
	300	(141.9, .509)	(100.5, .428)	342.9	(.973, .00163)
30	100	(44.9, .257)	(33.0, .187)	110.8	(.974, .00161)
	200	(89.3, .366)	(63.3, .308)	215.9	(.965, .00184)
	300	(133.6, .446)	(94.7, .375)	323.0	(.970, .00172)

Procedure (4.4) $(i = 0, 1, 2)$				
m	n^{**}	$(\overline{N}_i,\ S(\overline{N}_i))$	$\sum_{i=0}^{2} \overline{N}_i$	$(\overline{CP},\ S(\overline{CP}))$
10	100	(48.0, .352)	144.0	(.969, .00173)
	200	(94.9, .494)	284.8	(.964, .00187)
	300	(143.2, .610)	429.7	(.968, .00177)
15	100	(41.4, .297)	124.2	(.964, .00187)
	200	(83.0, .424)	249.0	(.961, .00194)
	300	(124.3, .510)	373.0	(.955, .00206)
20	100	(38.5, .264)	115.6	(.952, .00213)
	200	(76.9, .377)	230.6	(.952, .00213)
	300	(115.6, .459)	346.9	(.950, .00218)
30	100	(36.1, .219)	108.2	(.947, .00225)
	200	(71.0, .326)	212.9	(.947, .00225)
	300	(106.1, .397)	318.3	(.954, .00209)

We can observe from these tables that the proposed two-stage procedure (3.2)-(3.3) is inclined to classify π_0 exactly at a higher accurate than the two-stage procedure (4.4). The procedure (4.4) is inclined to undersample for π_0, and is inclined to oversample for π_1 and π_2. The proposed two-stage procedure (3.2)-(3.3) is inclined to require a smaller total sample sizes than the procedure

(4.4). Therefore, we conclude that the proposed two-stage procedure is more economical and we recommend to make use of it for the purpose of this problem positively.

ACKNOWLEDGMENTS

A part of this paper was completed during Aoshima's visit to University of Toronto. Research of the second author was supported by Grant-in-Aid for Encouragement of Young Scientists, the Ministry of Education, Science, Sports and Culture, Japan, under Contract Number 721-8035-09780215.

BIBLIOGRAPHY

[1] Aoshima, M. and Aoki, M. (1997). A note on fixed-size discriminant analysis for multivariate normal populations. *Dept. Math. Tech. Report 97-07-01, Syracuse University.*

[2] Aoshima, M. and Shimono, T. (1998). Fixed-size discrimination rule when two variances are equal, submitted.

[3] Bechhofer, R. E. and Turnbull, B. W. (1971). Optimal allocation of observations when comparing several treatment with a contorol, III, *Statistical Decision Theory and Related Topics, Vol. I* (Eds. Gupta and Yackel), Academic Press, 41-78.

[4] Das Gupta, S. (1974). Probability inequalities and errors in classification. *Ann. Statist.,* **2**, 751-762.

[5] Hochberg, Y. and Tamhane, A. C. (1987). *Multiple Comparison Procedures.* Wiley.

[6] Srivastava, M. S. (1973). A sequential approach to classification: Cost of not knowing the covariance matrix. *J. Multivariate Analysis.,* **3**, 173-183.

[7] Srivastava, M. S. and Khatri, C. G. (1979). *An Introduction to Multivariate Statistics.* North Holland.